

MIRTHIYA VIJAYARAGHAVAN

Data Scientist

9360387229 | ragavanmirthiya@gmail.com | [LinkedIn](#) | [Github](#)

Summary

Data science student with hands-on practical experience in developing predictive models, multimodal deep learning systems, and advanced analytics solutions. Skilled in **Python, Scikit-learn, TensorFlow, SQL, and visualization tools**. Having keen interest in the fields of machine learning, deep learning, and reinforcement learning. **Applied ML** techniques to real-world internships and academic research projects in healthcare, cybersecurity, and operational datasets. Excelled in identifying patterns in data, comprehending the meaning of models, and transforming complex information into useful outcomes.

Technical Skills

Programming Languages: Python, SQL

Machine Learning & AI: Supervised & Unsupervised learning, Model Training, Evaluation

Metrics, K-Means, DBSCAN, LSTM, Random Forest, Logistic Regression, Deep learning, Reinforcement Learning

Natural Language Processing: BERT, Transformers

Data Processing & Analysis: Pandas, NumPy, Statistical testing, Descriptive Statistics, Exploratory Data Analysis

Visualization & Monitoring: Matplotlib, Seaborn, Dashboards, ROC-AUC, precision, recall, F1-score

Tools & Platforms: Git, Kaggle, Jupyter Notebook, Microsoft Excel

Soft Skills: Problem-solving, Team management, Communication, Critical Thinking, Analytical Skills, Collaboration, Adaptability, Time Management.

Professional Experience

G Kuppaswamy Naidu Memorial Hospital, Coimbatore | Data Scientist Intern
April 2025 – May 2025

- Developed a machine learning model based on historical patient records for predicting outpatient revisit likelihood to enable early identification of high-risk patients.
- Carried out data cleaning, feature engineering, and exploratory analytics on OPD visit data that helped identify key predictors like visit frequency, category of diagnosis, and demographic factors.
- Built and tested several supervised ML algorithms (Logistic Regression, Random Forest, Gradient Boosting) to classify cases with revisit versus no-revisit in order to improve the prediction accuracy across models.
- Designed a data preprocessing pipeline for missing clinical information encoding categorical features as well as balancing the classes of revisits for better stability of the model
- Evaluated models by AUC precision-recall F1 score selection of the best classifier for deployment readiness

- Clinician-friendly insights were generated explaining model outputs using SHAP-based interpretability which supports decision-making about follow-up scheduling
- Worked with hospital IT and clinical staff to understand workflow needs and give actionable recommendations on reducing unnecessary patient revisits.

IBM SkillsBuild Program | Data Scientist Intern

September 2025

- Conducted an end-to-end exploratory data analysis (EDA) on employee demographic and job-related datasets to discover salary-driving factors like experience, job role, education level, and location.
- Created salary prediction models with supervised machine learning techniques (Linear Regression, Random Forest Regressor, Gradient Boosting), enhancing prediction accuracy through iterative tuning.
- Engineered relevant features, dealt with missing values, normalized numerical attributes, and encoded categorical variables for better model performance.
- Created multiple regression pipelines and compared them using RMSE, MAE, and R^2 metrics to find the most accurate salary prediction model.
- Visualized workforce patterns using Python (Pandas, Matplotlib, Seaborn), generating insights on compensation trends as well as employee distribution and role-based pay variation.
- Documented the complete modeling process from data preprocessing to EDA insights and evaluation results for clear reproducibility and presentation to mentors.
- Shared results using dashboards and charts that support data-driven recommendations for HR decisions about how to structure salaries.

Cognifyz | Data Analyst Intern

July 2025

- Conducted thorough Python-based EDA analysis on restaurant datasets to surface major trends in cuisines, pricing, ratings, and customer engagement.
- Developed effective visualizations (bar charts, histograms, scatter plots, geo-plots) to convey insights regarding restaurant locations, rating distributions, and price segmentation.
- Implemented K-Means clustering to identify geographic clusters and customer density hotspots based on longitude–latitude information.
- Assessed leading cuisines and restaurant chains through aggregated metrics like votes, average ratings, and cost for deriving actionable business insights.

Education

- **Int. M.Sc. Data Science**, Amrita Vishwa Vidyapeetham (2022–2027) **CGPA: 7.74**
- **Higher Secondary Education (Computer Math)**, RJ Mantra English School, ISC Board (2020–2022) **Grade: A (439/500)**
- **Secondary Education**, Sruthi Vidhyodhaya, ICSE Board (2019–2020) **Grade: A (429/500)**

Certifications

- IBM: Artificial Intelligence Fundamentals (SkillsBuild) Jul 2025
IBM: SkillsBuild Program July 2025
- Tata Forage Virtual Internship: Data Visualization – Empowering Business with Effective Insights Feb 2025 – Jun 2025
- Great Learning Academy: Statistics for Machine Learning Dec 2024
- Great Learning Academy: Excel for Beginners Dec 2024

Other Projects

Breast Cancer Prediction Using ML

Classification models were created using the features of a diagnostic tumor to separate malignant from benign cases with high accuracy, along with insights into feature importance for early detection of cancer.

Multi-Modal Retinal Disease Diagnosis

Developed a combined model that uses CNN-based image features along with organized patient information for better classification of retinal diseases, backed by visual justification through Grad-CAM heatmaps.

ESG-Driven Stock Prediction

Designed system for predicting stock movements that uses different types of information: market data, technical indicators, ESG reports, and sentiment/NER features based on natural language processing. These elements work together in one pipeline to allow deeper feature integration and show significantly better forecasting results than usual single-source models.